OPEN-SOURCE SOFTWARE-OPERATED CMOS CAMERA FOR REAL-TIME MAPPING

H. Gontran [a], J. Skaloud [a], N. Janvier [b]

[a] Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland – (herve.gontran, jan.skaloud)@epfl.ch
[b] Ecole Supérieure des Géomètres et Topographes (ESGT), Le Mans, France – nico.janvier@wanadoo.fr

ICWG V/1 – Integrated Systems for Mobile Mapping

KEY WORDS: Automation, Calibration, CMOS, Close Range, Mobile Mapping, Real-Time

ABSTRACT:

Accurate time registration, georeferencing, and instantaneous feature extraction are important challenges for real-time mapping systems. The Geodetic Engineering Laboratory (TOPO) at the EPFL has designed a mobile platform to automatically determine the geometry of the road centerline: the *Photobus*. Our system fully exploits the concept of direct georeferencing by combining navigation data with video sequences grabbed by a downward-pointing camera. This paper presents the realization of the authors' vision of real-time mapping that encompasses a mobile acquisition platform dedicated for a well-defined task. To sustain the high performance requirements of such a data-critical application, the use of custom-made software based on open-source libraries is a preferred approach of implementation. In this context, we investigate the suitability of using a close-to-nadiral CMOS camera to extract the road geometry with a reliable and almost latency-free methodology. We will first focus on the frame-grabbing process for which an open-source approach guarantees the versatile integration of the camera into the mapping platform. The second part of the paper discusses the process of extracting and georeferencing features of interest in an automatic manner. To open perspectives, we will investigate an extension of the surveying capabilities of the *Photobus* with stereovision.

1. INTRODUCTION

1.1 Mobile Mapping

The transport-telematics market witnesses considerable expansion, and the number of vehicle-navigation systems is continuously increasing. This success now causes the development of road-oriented cartographic databases that satisfy the Geographic Data File (GDF) standard. Only a close-range view allows road managers to assess the quality of their network, so that traditional methods for updating road databases often prove insufficient. Satellite-based remote sensing and aerial photogrammetry indeed offer reasonable costs; however, they are less suitable for our purpose due to low resolution or due to limited visibility of road objects that are masked by high buildings or trees. The early nineties experienced substantial progress in GPS/INS processing, followed by the market launch of affordable digital cameras with sufficient quality. This has enabled the acquisition of a significant part of the road data by vehicles equipped with navigation and videogrammetry tools (Ellum and El-Sheimy, 2002): land-based mobile-mapping systems were developed. However, such a technically-advanced method goes together with high investments in staff and hardware.

1.2 Real-time Mapping at the EPFL

To ensure direct quality control in the field, the information captured by a mobile-mapping system should be processed in real time. Such an approach may reduce the need for human intervention to driving the data-collecting platform while ensuring immediate quality control. The TOPO team at the EPFL has been investigating this topic through the road-geometry extraction from painted marks via the *Photobus* project (Gontran et al., 2004). This paper describes a promising real-time mapping strategy that involves a smart logarithmic CMOS camera operated by open-source software.

2. SMART CAMERA FOR OUTDOOR MAPPING

2.1 Logarithmic Pixels

The intra-scene brightness range of "typical" real-world environment covers five orders of magnitude, from 1 lux in shady areas to 10^5 lux in direct sunlight (Figure 1). Unfortunately, the charge-coupled devices (CCDs) and standard complementary metal-oxide semiconductor (CMOS) that dominate the image-sensor market today have a dynamic range of less than three orders of magnitude.

Figure 1. Road scene mixing shade and direct sunlight

Although these linear sensors can adapt over a high dynamic range by aperture adjustment or global control of integration time, saturated patches of black or white appear when capturing a scene of high dynamic range in a single picture. Yadid-Pecht (1999) reviews the different approaches to achieving an image sensor of high dynamic range, but most result in a low fill factor (the percentage of pixel area devoted to light collection) or large and expensive sensor elements. Only available for CMOS technology, logarithmic pixels represent a promising and low-cost alternative that can capture details in the bright and dark parts of a scene at the same time, thus approximating human vision. They can provide pixel values virtually at any moment in time, and nearly instantaneously (Scheffer et al., 1997). This renders them an ideal tool for capturing frames from a moving platform. Thus, logarithmic CMOS image sensors are particularly useful for a high dynamic range and high-speed imaging. Their drawback is an augmented sensitivity to fixed pattern noise (FPN), which is caused by a variation of device parameters, especially threshold voltages, from pixel to pixel, or column to column (Figure 2, left). This issue will be treated further below.

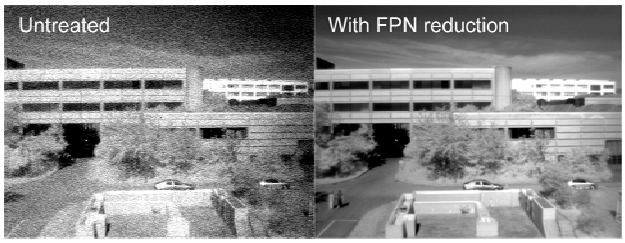

Figure 2. Images captured by a logarithmic CMOS image sensor, untreated or with FPN reduction

2.2 Ethercam: A logarithmic CMOS Image Sensor

We exploit the state-of-the-art CMOS imaging technology as the core of the *Ethercam* sensor that is embarked on the *Photobus*. The latter is a small-footprint vision system that remotely transmits processed or raw images to a host computer through the Ethernet connection. Consequently, the link to the computer hosting the mobile-mapping software is simple and does not require frame-grabber programming. Moreover, the camera's serial RS232 interface and three Transistor-Transistor-Logic (TTL) lines allow synchronizing image acquisition with GPS time.

The Linux-based operating system of the *Ethercam* handles the functions of image acquisition and digital preprocessing via an open-source application programming interface (API). This allows to digitally address the FPN issue directly on board by custom software. Such a method is based on reference noise values. These are obtained by taking a picture of a white sheet of paper under steady illumination and storing it inside an *Ethercam* flash memory module. This image represents the lowest order variation of pixel responses – the offset variation – and is subtracted from subsequent images that are captured (Joseph and Collins, 2001). Our practical tests reveal that this simple algorithm is sufficient to address the FPN issue. The quality of the correction process does not significantly degrade when the illumination of grabbed scenes departs from the uniform illumination used for calibration (Figure 2, right).

2.3 Synchronization of Frame Capture with GPS Time

Accurate time registration is a critical issue in direct georeferencing when the sequences are acquired from a moving platform. A millisecond-level synchronization is sufficient when mapping the close surroundings of the *Photobus* at a speed of 70 km/h (the vehicles moves 2 cm in 1 ms) with an accuracy of better than 10 cm. This justifies that all peripherals composing our mobile-mapping system can share TTL-compatible signals, i.e. synchronizing signals with a 10-µs resolution (Horowitz and Hill, 1989). Moreover, GPS receivers provide sub-microsecond timing under favorable satellite visibility. Thus, a highly satisfactory timing of picture acquisition can be obtained, provided that a TTL-enabled GPS receiver and one or several *Ethercams* are simultaneously triggered by adequate pulse edges. Open-source software, running on the host PC, conveniently intervenes in such a process. With this approach, a mere 10-line Python script is sufficient to make a computer serial port broadcast a duration-optimized square wave. This signal does not require being strictly cyclical; instead, it is only important to define the correct time of the frame capture. It is worth noting that in Figure 3, the serial voltage level is regulated and refined to achieve TTL compatibility. Besides, a custom-made splitting device distributes the modified signal among the GPS receiver and one or several *Ethercams* without introducing a temporal offset.

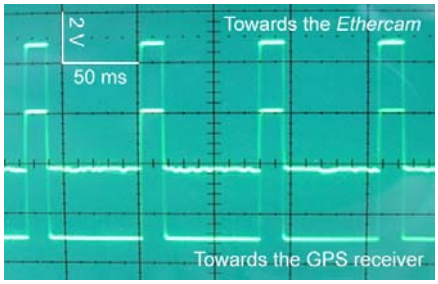

Figure 3. Simultaneity of TTL-compatible signals reaching a GPS receiver and the *Ethercam*

2.4 Storage of Captured Frames

A small-footprint device features an evident memory constraint, so that each grabbed frame is transmitted via TCP/IP to a host computer for storage and CPU-intensive processing. As Linux probably includes the most RFC[*]-compliant network stack available, this open-source operating system was the "natural" choice for operating the PC that collects the pictures. Its well-documented GNU GPL (General Public License) communication libraries greatly ease the implementation of the required client/server architecture (Gay, 2000). The server application, running on the *Ethercam*, first "pings" its client to verify the Internet connection. Then it waits for a TTL pulse to trigger the process of frame acquisition, FPN removal, and frame sending. The client application on PC allocates memory for incoming pictures, adds the appropriate picture header for subsequent recognition by our custom image processing software and stores them on a hard disk (Figure 4).

[*] Request for Comments (RFC) – a series of memoranda encompassing new research, innovations, and methodologies applicable to Internet technologies

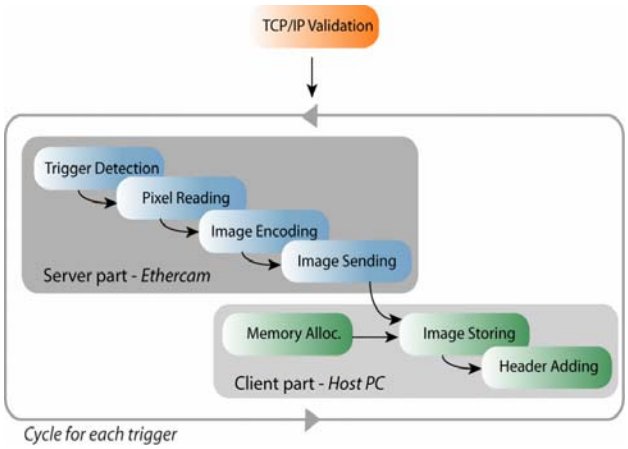

Figure 4. Client/Server interaction between *Ethercam* and PC during image acquisition

3. REAL-TIME GEOREFERENCING

3.1 Sensor-Supporting Structure

A rigid structure guarantees the stability of the relative placement and alignment of sensors, which is a prerequisite for applying the direct georeferencing methodology. Due to legal traffic restrictions, we designed a tee joint to be mounted on top of the vehicle, with a 3.4-meter-long stainless-steel section including two 0.6-meter-long cross-sections and a 1.65-meter-long aluminum cross-section taking into account the principles of materials science (Figure 5). The first section supports two GPS antennas with a separation of 2.2 m (in the along direction of the vehicle). Once connected to a dual-antenna RTK receiver, these allow computing a decimeter-level trajectory, as well as a degree-level azimuth under good satellite visibility. The second section can support:

- a downward-pointing *Ethercam* for the road-marking survey with a near 10-degree inclination that avoids imaging the vehicle side,
- a stereoscopic pair of forward-looking *Ethercams* to broaden the range of the *Photobus* applications with a downward inclination to obtain stereovision for photogrammetric exploitation of a roughly 10 m-by-10 m area that is situated of some 10 to 20 meters ahead of the vehicle.

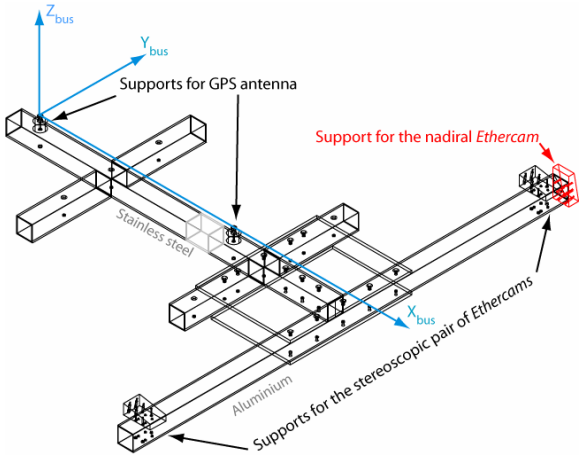

Figure 5. Sensor-supporting structure

3.2 Calibration of the Close-to-nadiral Ethercam

The camera calibration is relatively simple when considering the road as an inclined plane. It consists in the estimation of a suitable model for a close-to-nadiral *Ethercam* viewing the pavement from a height of approximately 2 meters. The goal is to find the parameters of exterior orientation (position and orientation relative to a world coordinate system), and parameters of interior orientation (principal point or image center, focal length and distortion coefficients of the camera) via an optimized algorithm. Tsai's camera model, a highly popular method among machine-vision researchers (Tsai, 1986), totally complies with our needs. It uses a two-stage technique to first compute the position and orientation and the interior orientation parameters of the camera afterwards.

Tsai's technique is based on the central perspective projection model and provides an estimate of the following parameters:

- f – focal length of an idealized pinhole camera
- k – first order radial lens distortion coefficient
- C_x, C_y – coordinates of the principal point
- S_x – scale factor accounting for uncertainty due to horizontal scan line resampling by the image sensor
- R_x, R_y, R_z – Euler angles for the rotation R matrix between the world coordinate system and the sensor frame defined below
- T_x, T_y, T_z – translation components for the origin difference T between the world coordinate system and the sensor frame

Figure 6 illustrates the geometry of the camera model. (x_w, y_w, z_w) are the coordinates of the object point P in the 3D world coordinate system defined by axes (X_w, Y_w, Z_w). (x_s, y_s, z_s) are the 3D metric coordinates of P in the 3D sensor coordinate system spanned by (X_s, Y_s, Z_s) that is centered at point O_s, the camera optical center, with the Z_s axis corresponding to the optical axis. (X, Y) are the image coordinate system centered at O_i (intersection of Z_s and the image plane) and parallel to the X_s and Y_s axes. (x_u, y_u) are the image coordinates of P if a perfect pinhole camera model is used. (x_d, y_d) are the actual image coordinates that differ from (x_u, y_u) due to lens distortion.

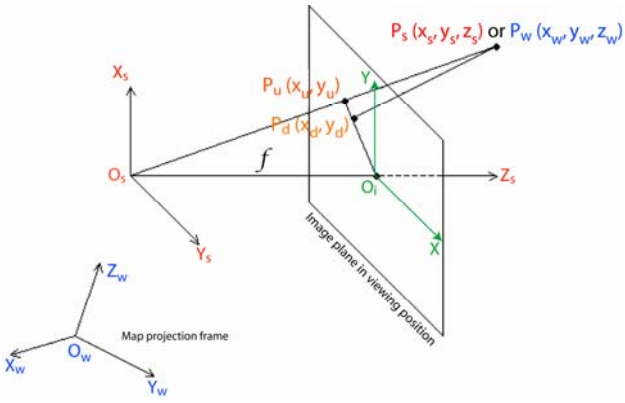

Figure 6. Tsai's camera model

The transformations involved in Tsai's model are depicted in the series of equations (1) where d_x and d_y are the distances between adjacent pixels in the X and Y directions respectively, and (x_f, y_f) defines the final pixel position. It is worth noting that Tsai introduces in his original model d'_x as d_x affected by a ratio between the number of sensor elements in the X direction

and the number of pixels in a line as sampled by the camera. This ratio is equal to 1 for the camera. The reader should consult Horn (2000) to obtain details about the determination of the 11 camera parameters.

$$
\begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T
$$
$$
\begin{bmatrix} x_u \\ y_u \end{bmatrix} = \frac{f}{z_s} \begin{bmatrix} x_s \\ y_s \end{bmatrix}
$$
$$
\begin{bmatrix} x_d \\ y_d \end{bmatrix} = \frac{1}{1+kr^2} \begin{bmatrix} x_u \\ y_u \end{bmatrix} \quad \text{where } r = \sqrt{x_d^2 + y_d^2} \tag{1}
$$
$$
\begin{bmatrix} x_f \\ y_f \\ 1 \end{bmatrix} = \begin{bmatrix} S_x / d_x & 0 & C_x \\ 0 & 1/d_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix}
$$

Tsai's algorithm is implemented in TLIB, a GNU-GPL C++ vision library written at the EPFL for real-time object tracking (Grange et al., 2003). Its freely accessible source code allows to include a driver for the *Ethercam*. Both the driver and the library were tested during an experiment that involved imaging of a 2 m-by-1.7 m calibration field by the close-to-nadiral *Ethercam*. A GPS-RTK survey provided the coordinates of all the circular calibration targets as a suitable reference, while their pixel coordinates were automatically given by a TLIB-based blob analysis (Figure 6). It is worth noting that the position of the structure supporting the GPS antennas was also determined during the RTK survey in order to define a bus-body frame.

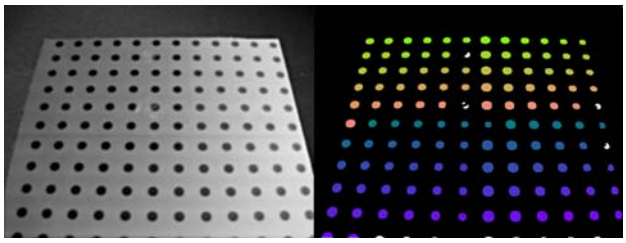

Figure 6. Automatic detection of calibration targets; on the right, colors validate the detection, white blobs were not detected.

Tsai's algorithm was applied using only the coordinate sets of the four corner targets of the field, as well as some others close to the image center. The quality of the obtained camera parameters was tested by using the rest of the target points as check points and comparing their world coordinates. As depicted in Figure 7, the camera model gives very satisfactory results: the mean 2D offset between true and computed targets corresponds to 2.0 mm and its RMS is 1.3 mm matching 0.3 pixels.

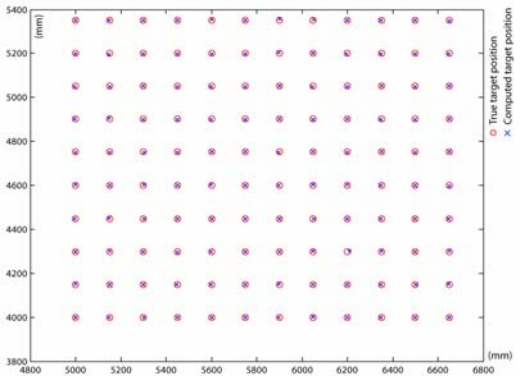

Figure 7. Comparison between Tsai-derived target positions and true target positions

3.3 Real-Time Georeferencing of Road Markings

Although less advanced than the robust detection approaches by Alvarez et al (2004) or Jiang and Quan (2005), our blob-analysis algorithm can easily be adapted to automate the detection of road markings and the subsequent extraction of their barycenter. The only elements to adapt are the binarization threshold and the topological constraints on the blobs. Figure 8 shows road scenes that are out of reach of CCD-sensors but can be successfully imaged by the *Ethercam* and processed by the TLIB-based software. The CPU-demanding TLIB library cannot indeed fit into the *Ethercam*, but provides results before a new picture needs to be processed. Consequently, this externalization of image processing still complies with real-time requirements, but we must admit that the TLIB responsiveness was not challenged by the *Ethercam* image transfer rate of only 3 frames per second.

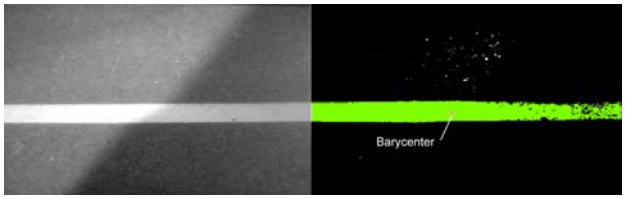

Figure 8. Barycenter extraction for a road mark under challenging illumination conditions

Once Tsai's calibration is performed, the transformations between the pixel- and sensor-coordinate systems and between world- and sensor-coordinates systems are known during the calibration step. Thereby, we define the body frame of the vehicle as follows:

- its origin is the phase center of the rear GPS antenna,
- its abscissa axis X_{bus} passes through the phase center of the front antenna,
- its Y_{bus} axis defined to be parallel to the plane spanned by (O, X_w, Y_w),
- its Z_{bus} axis is defined so that (X_{bus}, Y_{bus}, Z_{bus}) is an orthogonal right-handed coordinate system (see again Figure 5),

we can compute the transformation between the world- and the vehicle-coordinate frames. Afterwards, it is easy to define the 3D-Helmert transformation between sensor- and vehicle-coordinate systems. Consequently, pixel coordinates can be related to vehicle-frame coordinates and this relation is stable provided that the structure bearing the sensors is rigid.

On the other hand, the transformation between the bus frame and the national coordinate system changes as the vehicle moves. Its parameters are given by a dual-antenna GPS-RTK receiver for our first experiment involving the real-time extraction of the road axis. Consequently, translation elements derive from the position of the rear GPS antenna, and rotation is only azimuth- and pitched-based. We are planning a real-time GPS/INS integration to provide full sensor orientation and usability of the mapping platform under poor satellite visibility.

Our first tests involving the real-time algorithms show no apparent quality degradation when compared with the post-processing methodology for georeferencing, whose results can be seen in Gilliéron et al. (2005). Note that all the matrix computations are performed with an open source C++ library.

4. PERSPECTIVES

As an alternative to a close-to-nadiral camera, a forward-pointing stereoscopic pair of *Ethercams* enhances *Photobus*' surveying capabilities up to 20 m ahead of the mapping vehicle. In this context, the focus is on the suitable extraction of the road axis using photogrammetric plotting through open-source software (Janvier, 2005). The georeferencing process implies defining the transformation from the pixel coordinate system of each camera to the vehicle coordinate system. The steps involved are roughly the same as those described in section 3.3, although the pixel-to-world coordinate conversion is based on the collinearity equations. Moreover, such a close-range photogrammetry application requires a camera calibration model that takes into account several coefficients for radial and tangential lens distortion. The freely available camera calibration toolbox for Matlab (Bouguet, 2005) fulfills these needs; at the same time, it provides calibration parameters in the same range as those of the market-leading BINGO software (Kruck, 2003; see also Figure 9).

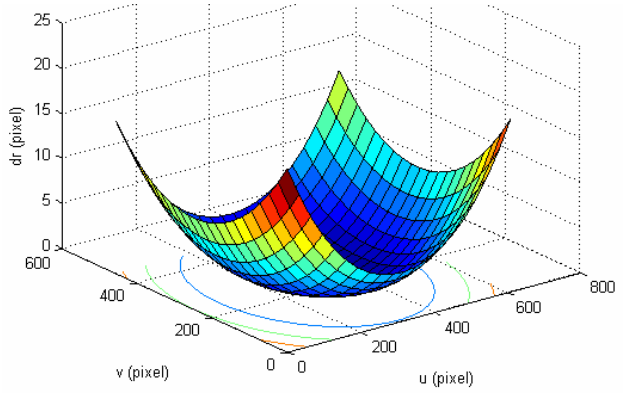

Figure 9. Distortion model for an *Ethercam* with Bouguet's camera calibration toolbox for Matlab

First efforts have been spent on bridging the gap between the manual and fully automated pixel-coordinate retrieval of homologous points. Therefore, we fused two open-source Matlab toolboxes (Kovesi, 2000; Torr, 2002) to implement a new tool that automatically matches image points. At calibration time, this tool first estimates the fundamental matrix *F* of the pair of *Ethercams* using a RANSAC (Random Sampling Consensus) process fed with homologous points that are collected by Harris' corner detector. Then, it iteratively applies the 8-point algorithm to refine the *F*-matrix elements, with a minimized Sampson distance as a stop criterion (Figure 10). Once *F* is optimized, the tool can be used for production and invites the user to select an image point from a stereo pair and exploits the now-defined epipolar geometry to reduce its search area in the corresponding image from a plane to a line. Finally, correlation based on the transitions in the pixel values is used to extract the matching point (Figure 11).

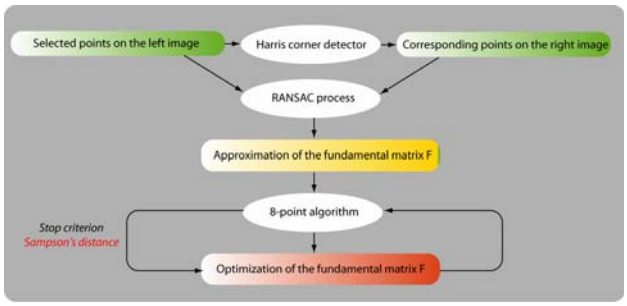

Figure 10. Definition of the epipolar geometry of the *Ethercam* stereopair (at calibration time)

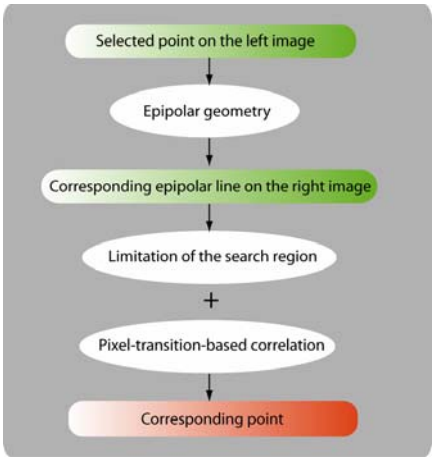

Figure 11. Exploitation of the epipolar geometry of the *Ethercam* stereopair (at production time)

A traditional restitution process using the inverse collinearity equations is currently implemented for determining the road axis geometry. First tests involving stereovision by an *Ethercam* pair show that simultaneous picture retrieval from both cameras needs enhancing. We suspect problems related to the hub that repeats incoming TCP/IP packets to all of its ports, which results in data collisions and sporadic frame loss. By definition, a switch supports simultaneous connections of computers that do not compete with each other for network bandwidth; its use should address data collisions.

5. CONCLUDING REMARKS

We explored two open-source-software-based techniques that offer promising results for real-time mapping of object situated on the road surface from a terrestrial vehicle. The first method – exploiting a close-to-nadiral camera for road-geometry determination – is fully implemented, although it needs extensive testing. A real-time road modeling by spline should be added to ensure immediate quality control of the automated procedure. The second method using stereoscopic vision extends the range of the *Photobus*' applications although it still

requires intensive efforts to reduce human intervention. However, the further development of freely available software libraries seems to be promising and should contribute to the automation of mapping process.

REFERENCES

Alvarez, L., Salgado, A., Sánchez, J., 2004. Robust Detection and Ordering of Ellipses on a Calibration Pattern. *Proceedings of the 7th International Conference on Pattern Recognition and Image Analysis*, St. Petersburg, Russia.

Bouguet, J.-Y, 2005. Camera Calibration Toolbox for Matlab®, http://www.vision.caltech.edu/bouguetj/calib_doc/

Ellum, C., and El-Sheimy, N., 2002. Land-based Integrated Systems for Mapping and GIS Applications. *Survey review*, 36 (283), pp. 323-339.

Gay, W., 2000. Linux Socket Programming by Example. *Que/Sams Publishing*. ISBN 0-7897-2241-0

Gilliéron, P.-Y., Gontran, H., Frank, J., 2005. High Quality Road Mapping for In-vehicle Navigation Safety. *Proceedings of the European Navigation Conference GNSS 2005*, Munich, Germany, 19-22 July.

Gontran, H., Skaloud, J., Gilliéron, P.-Y., 2004. Photobus: Towards Real-time Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, Vol. XXXV, Part B.

Grange, S., Fong, T., Baur, C., 2003. TLIB: a Real-Time Computer Vision Library for HCI. *Proceedings of the 7th Digital Image Computing: Techniques and Applications Conference*, Sydney, Australia, pp. 1017-1026.

Horn, B. K. P., 2000. Tsai's Camera Calibration Method Revisited. *Available from* http://people.csail.mit.edu/bkph/articles/Tsai_Revisited.pdf

Horowitz, P. and Hill, W., 1989. The Art of Electronics, 2nd Edition. *Cambridge University Press*, ISBN 0-521-37095-7.

Janvier, N., 2005. Identification des caractéristiques d'une route par stereovision mobile CMOS. http://www.esgt.cnam.fr/fr/tfe/memoires/2005/05_janvier_mem.pdf

Jiang, G., Quan, L., 2005. Detection of Concentric Circles for Camera Calibration. *Proceedings of the 10th International Conference on Computer Vision*, Beijing, China, pp. 333-340.

Joseph, D. and Collins, S., 2001. Modeling, Calibration and Correction of Nonlinear Illumination-dependent Fixed Pattern Noise in Logarithmic CMOS Image Sensors. *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hungary, Vol. II, pp. 1296-1301.

Kovesi, P. D., 2000. Matlab and Octave Functions for Computer Vision and Image Processing. *Available from* http://www.csse.uwa.edu.au/~pk/research/matlabfns/

Kruck, E., 2003. BINGO Bundle Adjustment for Engineering Applications, Version 5.0. *Software Reference Manual*.

Combined IMU and sensor calibration with BINGO F. i*Integrated Sensor Orientation, Proceedings of the OEEPE Workshop*, Hannover, Germany, CD-ROM.

Scheffer, D., Dierickx, B., Meynants, G., 1997. Random Addressable 2048×2048 Active Pixel Image Sensor. *IEEE Transactions on Electron Devices*, Vol. XLIV, no. 10, pp. 1716-1720.

Torr, P., 2002. A Structure and Motion Toolkit in Matlab. *Available from* http://cms.brookes.ac.uk/staff/PhilipTorr/

Tsai, R., 1986. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, USA, pp. 364-374.

Yadid-Pecht, O., 1999. Wide-dynamic-range Sensors. *Optical Engineering*, Vol. XXXVIII, no. 10, pp. 1650-1660.